

## Vigilancia Tecnológica e Inteligencia Competitiva basada en Técnicas de Minería de la Web

Karina Eckert<sup>1</sup>, Victor Alvarenga<sup>1</sup>, Matias Barboza<sup>1</sup>, Leandro Witzke<sup>1</sup>, Luis Airaldi<sup>2</sup>

<sup>1</sup>Univer. Gastón Dachary. Av. López y Planes 6519, Posadas, Misiones.

<sup>2</sup>Univ. Tecnológica Nacional, Facultad Regional Resistencia. French 414, Resistencia, Chaco.  
{karinaeck, matias.mbz, leanwitzke, luislezcair}@gmail.com, alva\_victor@hotmail.com

**Abstract.** Tomar decisiones estratégicas es un proceso complejo que requiere información fiable y actualizada. Por ello es necesario contar con herramientas que faciliten la administración de la información. La Vigilancia Tecnológica y la Inteligencia Competitiva (VT e IC) son dos disciplinas que tienen como objetivo la obtención de información precisa y actualizada. Claramente, la Web es la fuente más grande e importante de información, pero su desestructuración y desorganización requiere de herramientas que ayuden a administrarla. En este trabajo se presenta un modelo para hacer VT e IC utilizando técnicas de Minería de Web. Se describen los módulos del sistema desarrollado y las primeras pruebas realizadas en el ámbito del Clúster del Té de Misiones.

**Keywords:** minería de Web, vigilancia tecnológica e inteligencia competitiva, toma de decisiones.

### 1 Introducción

La Toma decisiones (TD) estratégicas requiere del análisis en detalle de una cantidad considerable de información; la cual no se dispone en su totalidad en el momento preciso. Si bien la falta de información para la TD es habitual en las organizaciones, las del área productiva del Té de Misiones son de las más afectadas por diversas razones: los métodos tradicionales de trabajo, la poca capacidad de inversión de las PyMEs, el poco impacto internacional de la producción, etc. Por ello se ha creado el Clúster del Té de Misiones (PyMEs, cooperativas, organismos del estado y universidades), que intenta encontrar soluciones a problemas comunes de sus integrantes. En base a esta unión, se afronta el desafío actual de contar con nuevas formas de competencia, poner a prueba en grado crítico a las organizaciones para posicionarse y articularse con otros actores del entorno y promover el dominio y ejercicio de una capacidad continua de monitoreo o vigilancia del entorno. Para ello, existen dos actividades muy utilizadas para analizar las variaciones del entorno: la Vigilancia Tecnológica (VT) y la Inteligencia Competitiva (IC). Ambas tienen el objetivo de alertar cambios en las tendencias de la tecnología y el comportamiento del mercado. La VT engloba los métodos que intentan anticipar y entender la dirección potencial, las características y efectos de los cambios tecnológicos, especialmente en lo referido a invención, innovación y uso [1]. En la IC intervienen los métodos éticos y legales que tienen como

fin, obtener información valiosa sobre los competidores [2]. Las dos estrategias de captación de datos deben integrarse a los Sistemas de Soporte de Decisiones (SSD), los cuales proveen una serie de técnicas y herramientas que sirven para asistir al decisor durante el proceso de TD [3][4][5].

Se estima que un 80% de la información estratégica no se encuentra estructurada y organizada, donde gran parte de ella está en Internet. La Minería Web (MW) es una metodología de recuperación de información que permite procesar y capturar información útil de páginas Web y documentos en Internet que pueden servir para llevar a cabo procesos de VT e IC, contribuyendo en una de sus etapas principales que es la de búsqueda y recolección de información para su posterior análisis y tratamiento. Básicamente los algoritmos de MW usan técnicas de Minería de Datos para descubrir y extraer información de los documentos y recursos disponibles de la Web [6]. La MW puede dividirse en tres categorías: MW del contenido Web; MW de la estructura de la Web; y MW del uso de la Web [7]. De estas tres, la más relevante para la VT e IC es la primera, la cual abarca el descubrimiento de recursos, categorización y clustering de documentos y extracción de información de páginas Web [8][9][10].

Este artículo describe un prototipo de software basado en técnicas de MW que permita hacer VT e IC en el ámbito del Clúster del Té en Misiones. De esta manera el documento se estructura de la siguiente manera: en la Sección 2 se exponen los principales ejes temáticos del trabajo; en la Sección 3 se describe el modelo desarrollado; los resultados obtenidos se muestran en la Sección 4 y, finalmente, en la Sección 5 se mencionan las conclusiones obtenidas y trabajos futuros.

## **2 Conceptos preliminares**

### **2.1 Vigilancia Tecnológica e Inteligencia Competitiva**

Con la globalización de los mercados y el desarrollo de la era digital, el valor que las empresas atribuyen a la información ha ido cambiando y hoy es más estratégica que nunca. Esto ha hecho que disciplinas como la VT e IC se hayan convertido en un pilar esencial para crear nuevos productos o servicios, para definir estrategias de comercialización, potenciar las capacidades de la organización, mejorar el servicio a los clientes, etc. [11]. Existen muchos métodos de VT los cuales pueden agruparse en las siguientes clases [12][13]: análisis de tendencias; opinión de los expertos [14][15]; métodos de monitoreo e inteligencia [16]; métodos estadísticos; modelado de escenarios; y métodos de modelado y simulación [17].

La IC es una disciplina que se encarga de la obtención, análisis, interpretación y difusión de información del entorno competitivo en el que se mueven las empresas, a través de un proceso sistemático y ético, indispensable para la TD [11][18]. Existen tres etapas [19]: recolección de información, extracción de información y contextualización de la información. Los objetivos del uso de un sistema de información de IC destacan tres aspectos: mejorar la competitividad de la empresa; predecir, con un alto nivel de confianza, la evolución del entorno y proveer un buen soporte para el proceso de decisión estratégico [20].

## 2.2 Minería de la Web

La cobertura de la información en la Web es muy amplia y diversa; existen datos de todo tipo: tablas estructuradas y semi-estructuradas, páginas Web, textos no estructurados y archivos multimedia. Esto demuestra que existe una amplia heterogeneidad de problemas de índole multi-disciplinaria a ser resueltos con el fin de explotar satisfactoriamente la información Web como recursos valiosos para las empresas. WM se plantea como el conjunto de técnicas que permiten abarcar la problemática de explotación de la información que se encuentra en la Web [21].

El proceso de MW puede ser definido como “el proceso global de descubrir información o conocimiento potencialmente útil y previamente desconocido a través de los datos de la Web” [22]. La MW tiene como objetivo descubrir información útil o el conocimiento tanto del contenido de documentos Web, como también de la estructura de hipervínculos Web y los datos de uso [23].

Una vez que se hallan sitios Web con contenido importante, el sistema debe ser capaz de obtener todo lo relevante utilizando mecanismos de extracción de contenido Web (Web Scraping), y al mismo tiempo adjuntar información de indexación a los registros para mantener una estructura que facilite futuros accesos a los recursos Web de una manera más directa, sin la necesidad de analizar ni ponderar contenido de una página que ya ha sido considerado relevante [21]. Para guiar la navegación de enlaces y localizar de manera eficiente las páginas de destino de gran relevancia, se utilizan técnicas de Web Crawling; donde, cada página que es escaneada es dada a un cliente que guarda las páginas, crea un índice para las páginas o guarda y analiza los contenidos de las páginas [23].

## 3 Modelo desarrollado

En base a la visión de VT e IC se propone un modelo que cumpla dos objetivos: el primero generar un proceso para dar soporte a la recopilación de los requerimientos de usuario y el segundo generar un proceso de búsqueda continua de recursos en la web. Ambos procesos trabajan coordinadamente en un mismo flujo interactuando constantemente. La idea del primer proceso es orientar al usuario en el armado de varias claves que serán tomadas como punto de partida de la búsqueda y la del segundo es proveer a medida al usuario de recursos que se vayan encontrando. El esquema general del sistema puede observarse en la Fig. 1. El sistema desarrollado posee dos componentes principales el Módulo de Recopilación de Requerimientos de usuario y el Módulo de MW (Fig. 1). Ambos interactúan utilizando un servicio Web con el que se administra el proceso de comunicación. El usuario comienza interactuando con el Módulo de Recopilación de Requerimientos que, mediante una serie de preguntas, lo guía en la construcción de claves de búsqueda. Esas claves se envían al Módulo de MW quien consulta a las interfaces públicas de cuatro buscadores: Google, Bing, Intelligo (un buscador de patentes) y Msmx Excite (un metabuscador). Las primeras diez URLs de cada uno son categorizadas y ordenadas en una sola lista en base a los requerimientos de usuario (de esa lista se quitan las URLs repetidas). A partir de este punto el Módulo de MW comienza un proceso continuo de inspección de enlaces y

recursos tomando como base cada una de las URLs de la lista. A medida que se encuentran recursos que se consideran relevantes para el usuario, se los coloca en un directorio junto con los metadatos del recurso. Por las características de este proceso continuo (lo que genera un gran volumen de información), se seleccionan solamente cincuenta recursos que se le muestran al usuario para que los valore.

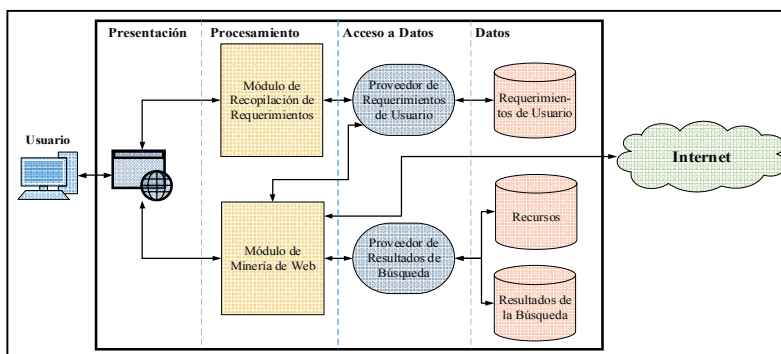


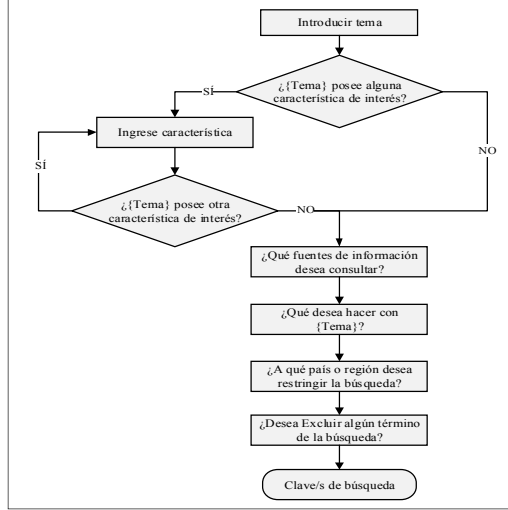
Fig. 1. Esquema general del modelo desarrollado.

### 3.1 Módulo de Recopilación de Requerimientos

Comúnmente los usuarios tienen dificultades para armar una clave de búsqueda que represente sus requerimientos de información y, por lo general, lo que hacen es introducir una serie de palabras relacionadas con la temática de búsqueda. Si bien estas claves devuelven recursos que pueden ser útiles, no explotan todo el potencial de los buscadores debido a que, generalmente, no se utilizan conectores lógicos ni símbolos para demarcar partes importantes de la clave.

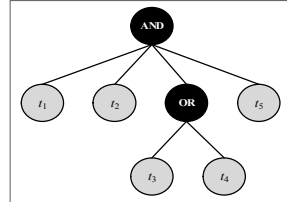
En el modelo propuesto, se utiliza el mecanismo de preguntas guiadas para orientar al usuario en la construcción de las claves de búsqueda descrito en [24]. Inicialmente el usuario determina el tema principal y luego continúa con un proceso de refinamiento mediante preguntas estándares (Fig. 2). Debido a que las claves de búsqueda generadas se usan sobre las mismas interfaces públicas de los buscadores, se requiere utilizar los operadores lógicos: OR, AND y NOT [25][26]. Cabe aclarar que si se desea armar una estructura lógica donde una operación tenga mayor relevancia que otra, se debe estructurar de manera tal que pueda obtenerse un resultado similar a utilizar paréntesis, pero sin indicarlos explícitamente. Otra característica importante es que los navegadores otorgan a los primeros términos de la clave mayor importancia mientras que las siguientes claves refinan el proceso de exploración. Una vez finalizado la recolección de requerimientos, se deben relacionar las respuestas a través de los dos operadores lógicos básicos definidos en el sistema: AND y OR. Estas relaciones se utilizan posteriormente para construir una estructura de árbol donde cada nodo hoja contiene la respuesta provista por el usuario y cada nodo padre contiene el operador lógico que relaciona a estas respuestas entre sí Fig. 3. La secuencia de preguntas se establece por medio de un atributo de relación entre ellas y luego se las pondera [24]. Aquellas preguntas que tengan una ponderación mayor, serán por lo tanto, más importantes y sus respuestas ocuparán los primeros lugares en el conjunto de claves

de búsqueda generadas. Seguidamente, se realiza el proceso de creación del árbol teniendo en cuenta el operador lógico que relaciona a cada pregunta  $P_i$  de la lista ordenada con la próxima pregunta  $P_{i+1}$ .



**Fig. 2.** Proceso de obtención de Requerimientos.

Como resultado, se obtiene una estructura de árbol como se muestra en la Fig. 3. Una respuesta puede estar compuesta por un término o un conjunto de términos separados por espacios. Estos términos se denotan con  $t_1, t_2, \dots, t_n$ .



**Fig. 3.** Árbol de respuestas para un esquema genérico de preguntas.

Finalmente, la clave de búsqueda se define como una cadena de caracteres compuesta por los términos o palabras importantes provistas por el usuario, enlazadas entre sí mediante los operadores lógicos predefinidos. Se han implementado dos opciones para la generación de las claves. La primera construye una clave única que contiene todos los términos proporcionados por el usuario, y la segunda separa cada opción de una respuesta con múltiples opciones en claves diferentes pensadas para ser utilizadas en conjunto en el mismo proceso de búsqueda [24]. De la primera opción se obtiene la siguiente clave sobre el árbol de la Fig. 3:

$$C = t_1 \cdot t_2 \cdot (t_3 + t_4) \cdot t_5 \quad (1)$$

De la segunda opción se obtienen dos claves sobre el mismo árbol:

$$C_1 = t_1 \cdot t_2 \cdot t_3 \cdot t_5 \quad (2)$$

$$C_2 = t_1 \cdot t_2 \cdot t_4 \cdot t_5 \quad (3)$$

esta opción separa los nodos hoja de un operador de disyunción en claves distintas (Ecs. 2 y 3). Por lo tanto, se obtienen  $N$  claves, donde  $N$  es el producto de la cantidad de nodos hoja subordinados a cada nodo OR considerando el árbol completo (esto provee un mecanismo más adecuado para el inicio del proceso de búsqueda).

### 3.2 Módulo de Minería de Web

El proceso de minería se inicia cuando el sub-módulo llamado Web Miner (Fig. 4) recibe el conjunto de claves generadas por el Módulo de Recopilación de Requerimientos. Estas claves son enviadas al sub-módulo de Búsqueda que se encarga de obtener URLs proporcionadas por las interfaces públicas de los cuatro motores de búsquedas utilizados (los resultados obtenidos por Google y Bing se obtienen mediante sus APIs, en tanto para Msmlx Excite e Intelligo se obtienen los resultados simulando un procedimiento de búsqueda manual). A partir del conjunto de URLs obtenidas, se genera una lista única filtrada de acuerdo a la posición que ocupan en dichos buscadores y la relación con los requerimientos establecidos por el usuario. Por ejemplo, las URLs de dominio académico (".edu") tendrán más peso si el usuario determinó que quiere artículos o publicaciones relacionadas con el tema de su búsqueda; en cambio, el mayor peso lo tendrán los dominios ".com" si lo que pretende es comprar o vender. Con esta lista ordenada, el sub-módulo de Web Crawler se encarga de comenzar el proceso de descubrimiento. Este proceso es continuo e iterativo a partir de dicha lista.

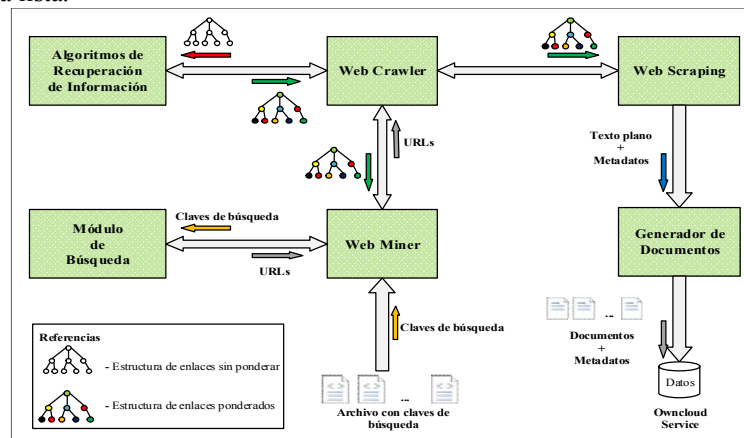


Fig. 4. Arquitectura básica de la aplicación.

Por cada URL se genera un grafo de  $n$  URLs descubiertas. Las aristas corresponden a las relaciones entre las URLs. El grafo es enviado al sub-módulo de Algoritmos de Recuperación de Información para establecer una puntuación a cada URL del grafo con el objetivo de obtener una lista ordenada según la relevancia del contenido. El grafo ponderado es enviado al sub-módulo Web Scraping que se encarga de descargar el contenido de las URLs más relevantes y generar los metadatos que contiene información sobre la ponderación realizada y la URL del documento. Esta información se envía al sub-módulo Generador de Documentos que crea los documentos dependiendo de la extensión (.pdf, .html, .asp, .php, etc.) y los asocia con los metadatos para

almacenarlo en un directorio que será sincronizado en todas las ubicaciones que requieran la información (Owncloud Service). Luego, el sub-módulo Web Crawler obtiene la siguiente URL de la lista y se reinicia el proceso.

Cabe destacar que el sub-módulo de Algoritmos de Recuperación de Información no solamente pondera las URLs del grafo recibido en cada iteración, sino que además vuelve a ponderar las URLs de los grafos que fueron procesadas anteriormente. El proceso vuelve a reajustar el peso de todas las URLs para generar una nueva lista de documentos ordenados por la relevancia del contenido. El proceso del sub-módulo Web Crawler finaliza cuando el usuario decide parar con la búsqueda de documentos o cuando no existen más URLs por explorar.

## 4 Pruebas y Resultados

Con el objetivo de verificar el funcionamiento del sistema desarrollado se han realizado tres pruebas, las dos primeras relacionadas con la búsqueda de información sobre el proceso productivo del té en Argentina, la exportación del mismo, relacionadas con el ámbito de aplicación del sistema; y la tercera, sobre las técnicas de machine learning para hacer análisis de datos para mostrar la extensibilidad del uso del sistema a otros dominios de aplicación. Los resultados que se presentan son los que se obtuvieron luego de 12 horas de iniciado el proceso.

### 4.1 Prueba 1: Proceso productivo del té

Esta primera prueba consistió en obtener recursos web respecto al proceso de producción del té. Para ello se utilizó el módulo de recopilación de requerimientos con las siguientes consignas: *proceso productivo del té, optimización, patentes, artículo, cita, reducción de costos, automatización y maquinaria, Argentina*, excluyendo *publicidad*. Puede observarse en la Tabla 1 que la gran mayoría de recursos obtenidos son de sitios académicos, especialmente presentaciones o slides que explican el proceso de producción del té.

**Tabla 1.** Porcentajes de sitios y recursos obtenidos para proceso productivo del té.

Tipo de sitios	Porcentaje obtenido	Tipo de recurso	Porcentaje obtenido
Comerciales	05%	Artículos de revista	24%
Gubernamentales	24%	Patentes	50%
Académicos, educativos	60%	Slides	10%
Otros	11%	Información Comercial	16%

Otro aspecto interesante es que los sitios gubernamentales de la lista corresponden en su gran mayoría al Instituto Nacional de Tecnología Agropecuaria (INTA) de Argentina donde se explican otros procesos productivos como el de la yerba mate o el algodón y, del mismo ámbito del INTA, también se ha encontrado en la lista un sitio de maquinaria agrícola y valor agregado a productos agrícolas.

### 4.2 Prueba 2: Exportación de té

En la segunda prueba se han utilizado las siguientes consignas para buscar información sobre exportación de té: *argentinian tea exportation, international trade,*

*traditional tea* y *potential markets*, excluyendo *advertising*. Los resultados se de los recursos obtenidos se pueden apreciar en la Tabla 2.

**Tabla 2.** Porcentajes de sitios y recursos obtenidos para exportación de té.

Tipo de sitios	Porcentaje obtenido	Tipo de recurso	Porcentaje obtenido
Comerciales	88%	Artículos de revista	06%
Gubernamentales	04%	Patentes	00%
Académicos, educativos	08%	Slides	00%
Otros	00%	Información Comercial	94%

El primer aspecto que se destaca en los resultados de la Tabla 2 es el elevado porcentaje de sitios comerciales encontrados. Entre ellos hay una gran variedad de dominios que incluyen información estadística de exportación de productos (no solamente té) de los distintos países. Uno de los sitios de la lista (theteadetective.com) aparece varias veces con distintas páginas (TeasOfArgentina, TeasOfTheWorld, BlendedScentedTeasDiffer, IndiaTeaGrowsAssam, etc), al revisar el enlace se descubrió que esta página tiene gran cantidad de información sobre los distintos aspectos acerca de la historia, producción, elaboración y comercialización del té.

#### 4.3 Prueba 3: Machine Learning para análisis de datos

Esta prueba tuvo como objetivo revisar la potencial extensión del uso del sistema a otros dominios de aplicación. En este caso se optó por buscar información sobre algoritmos de machine learning que se utilizan para análisis de datos. Para ello se trabajó con las siguientes consignas: *data analytics*, *big data*, *paper*, *cite*, *trends* y *statistics* excluyendo *advertising*. En este caso se intentó realizar una búsqueda de recursos netamente académicos y los resultados pueden observarse en la Tabla 3.

**Tabla 3.** Porcentajes de sitios y recursos obtenidos para machine learning.

Tipo de sitios	Porcentaje obtenido	Tipo de recurso	Porcentaje obtenido
Comerciales	00%	Artículos de revista	86%
Gubernamentales	00%	Patentes	00%
Académicos, educativos	92%	Slides	14%
Otros	08%	Información Comercial	00%

Los resultados mostrados en la Tabla 3 indican que casi todos los sitios son académicos. Entre ellos se destacan enlaces a artículos de distintas editoriales como Springer, Elsevier y sitios de manejo de información académica como Mendeley. El resto de los sitios encontrados contienen también información académica, pero en forma de blogs personales y de organizaciones como el caso de <http://squad.compmath.fr/> o <http://deeplearning4j.org/>, ambas orientadas al desarrollo de algoritmos de machine learning bajo distintos lenguajes de programación. También, como punto destacable, entre los resultados aparece el sitio del ICML 2015, que corresponde uno de los congresos sobre machine learning más prestigiosos. Finalmente, es de destacar que en esta prueba se ha destacado la gran cantidad de resultados potencialmente utilizables respecto a la búsqueda planteada.



## 5 Conclusiones y Trabajos Futuros

En este artículo se ha presentado un sistema que sirve de herramienta para actividades de Vigilancia Tecnológica e Inteligencia Competitiva. El sistema cuenta con dos módulos principales, el primero sirve para guiar al usuario en la recopilación de requerimientos para el proceso de búsqueda generando claves de búsquedas más completas que las que habitualmente se utilizan en los buscadores. El segundo módulo se encarga de la minería de Web, explorando enlaces a partir de URLs devueltas por las interfaces públicas de los buscadores más utilizados.

Un aspecto de relevancia del modelo propuesto es que no hace una búsqueda tradicional de consulta-respuesta, sino que los resultados que se obtienen son refinados continuamente y categorizados en función a los requerimientos de usuario. Esto significa que, una vez iniciado el proceso de búsqueda por parte del usuario, el mismo sigue ejecutando obteniendo nuevos resultados hasta que el usuario decida finalizarlo.

Si bien el análisis realizado en este artículo ha sido cuantitativo debido a que se han presentado las primeras experiencias realizadas con el sistema, las pruebas muestran que el sistema encuentra información acorde al tipo de respuesta esperada. En este sentido, las tres pruebas tenían objetivos distintos y los recursos obtenidos por el sistema coinciden con dichos objetivos.

Cabe mencionar que si bien se ha comenzado a probar el sistema en base a los requerimientos del Clúster del Té de Misiones como dominio de aplicación, las pruebas realizadas muestran que su uso puede ser extendido a otros dominios. Cabe mencionar que si bien se ha comenzado a probar el sistema en base a los requerimientos del Clúster del Té de Misiones como dominio de aplicación, las pruebas realizadas muestran que su uso puede ser extendido a otros dominios.

Actualmente se encuentra en desarrollo el proceso de ajuste de los algoritmos y el tratamiento e indexación de resultados. El primero permitirá mejorar la performance del sistema y el segundo facilitará la visualización de los resultados y el feedback del usuario. Además, se está trabajando en la aplicación de métricas de evaluación (“precision” y “recall”) para medir la recuperación de información.

### Agradecimientos

A los miembros del Clúster del té por facilitar el dominio de aplicación. Este trabajo forma parte de los proyectos “Modelos de Análisis de Información para la Toma de Decisiones Estratégicas” (UGD) y “Análisis de Información en Grandes Volúmenes de Datos Orientado al Proceso de Toma de Decisiones Estratégicas” (cód: UTN4058).

### Referencias

- [1] A. Firat, W. Woon, and S. Madnick, “Technological Forecasting – A Review,” presented at the Working Paper CISL# 2008-15, Cambridge, 2008.
- [2] R. G. Vedder, M. T. Vanecek, C. Guynes, and J. Cappel, “CEO and CIO Perspectives on Competitive Intelligence,” *Commun. ACM*, vol. 42, no. 8, 1999.
- [3] J. P. Shim and others authors; “Past, Present, and Future of Decision Support Technology,” *Decis. Support Syst.*, vol. 33, no. 2, pp. 111–126, 2002.
- [4] C. W. Holsapple, “Decision Support Systems,” *Encycl. Inf. Syst.*, vol. 1, pp. 551–565, 2003.

- [5] D. J. Power and R. Sharda, "Decision support systems," Springer Handbook of Automation, pp. 1539–1548, 2009.
- [6] O. Etzioni, "The World Wide WEB: Quagmire or gold mine," Commun. ACM, vol. 39, no.11, pp. 65–68, 1996.
- [7] R. Kosala and H. Blokeel, "WEB mining research: A survey," ACM SIGKDD Explor., vol. 2, no. 1, pp. 1–15, 2000.
- [8] S. Chakrabarti, M. van der Berg, and B. Dom, "Focused crawling: A new approach to topic-specific WEB re-source discovery," in In Proceedings of the 8th international conference on World Wide WEB (WWW8), pp. 545–562, 1999.
- [9] T. Kohonen, and others authors "Selforganization of a massive document collection," IEEE Trans. Neural Netw., vol. 11, no. 3, pp. 574–585, 2000.
- [10] C.H. Chang, C.N. Hsu, and S.C. Lui; "Automatic information extraction from semi-structured WEB pages by pattern discovery," Decis. Support Syst., vol. 35, pp. 129–147, 2003.
- [11] R. Barainka; "Modelos de Vigilancia Tecnológica e Inteligencia Competitiva". Servicio Zaintek de BAI. 2006.
- [12] V. Coates and others authors; On the future of technological forecasting, Technol. Forecast. Soc. Change, vol. 67, no. 1, pp.1-17, 2001.
- [13] T.J.Gordon and J.C. Glenn, Futures research methodology, Version 2.0 Millennium Project of the American Council for the United Nations University, 2003.
- [14] F. Parente and others authors; An examination of factors contributing to Delphi accuracy, J. Forecast; vol. 3, pp.173 – 182, 1984.
- [15] R. Levary and D. Han; Choosing a technological forecasting method. Industrial Management, vol.37, no. 1, pp.14, 1995.
- [16] S. Madnick and W.L. Woon; Technology Forecasting Using Data Mining and Semantics, MIT/MIST Collaborative Research, 2009.
- [17] T.J. Gordon; "A simple agent model of an epidemic" Technological Forecasting & Social Change, vol. 70, pp. 397-417, 2003.
- [18] Escorsa Pere; "INTEC, la Inteligencia Competitiva. Factor clave para la toma de decisiones estratégicas en las organizaciones", Plan Regional de Ciencia y Tecnología de la Comunidad de Madrid, 2008.
- [19] T. Hiltbrand; "Learning Competitive Intelligence From a Bunch of Screwballs", Business Intelligence Journal, vol: 15, no: 4, 2010.
- [20] I. Popa Anica and G. Cucui, "A Framework for Enhancing Competitive Intelligence Capabilities using Decision Support System based on Web Mining Techniques", Int. J. of Computers, Communications & Control, vol. 4, no. 4, pp. 326-334, 2009.
- [21] L. Bing; "Web Data Mining, Exploring Hyperlinks, Contents, and Usage Data", First Edition. University of Illinois, Chicago. 2007.
- [22] V. Heughes Jeria Escobar; "Minería Web de Uso y Perfiles de Usuario: Aplicaciones con Lógica Difusa", Universidad de Granada. 2007.
- [23] F. Menczer, G. Pant and P. Srinivasan; "Topical Web Crawlers: Evaluating Adaptive Algorithms". ACM Transactions on Internet Technology, vol. 5. 2004.
- [24] L. Lezcano and others authors; "Modelo de Recopilación de Requerimientos para Vigilancia Tecnológica e Inteligencia Competitiva (VTeIC)" 3er CONAIISI | 19 y 20 de Noviembre 2015. UTN, Facultad Regional Buenos Aires.
- [25] Support.google.com, 'Search operators - Search Help', Disponible en: <https://support.google.com/websearch/answer/2466433>, [último acceso: 05/07/2016].
- [26] Msdn.microsoft.com, 'Advanced Operator Reference', Disponible en: <https://msdn.microsoft.com/en-us/library/ff795620.aspx>, [último acceso: 06/07/2016].